Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: www.tandfonline.com/journals/lecr20

Flexible Threshold Models for Modelling Interest Rate Volatility

Petros Dellaportas, David G. T. Denison & Chris Holmes

To cite this article: Petros Dellaportas, David G. T. Denison & Chris Holmes (2007) Flexible Threshold Models for Modelling Interest Rate Volatility, Econometric Reviews, 26:2-4, 419-437, DOI: 10.1080/07474930701220600

To link to this article: https://doi.org/10.1080/07474930701220600

Published online: 04 May 2007.

Econometric Reviews, 26(2–4):419–437, 2007

ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474930701220600

Taylor & Francis
Taylor & Francis Group

FLEXIBLE THRESHOLD MODELS FOR MODELLING INTEREST RATE VOLATILITY

Petros Dellaportas □ *Department of Statistics, Athens University of Economics and Business, Athens, Greece*

David G. T. Denison □ *London, UK*

Chris Holmes □ *Department of Statistics, University of Oxford, Oxford, UK*

□ *This paper focuses on interest rate models with regime switching and extends previous nonlinear threshold models by relaxing the assumption of a fixed number of regimes. Instead we suggest automatic model determination through Bayesian inference via the reversible jump Markov Chain Monte Carlo (MCMC) algorithm. Moreover, we allow the thresholds in the volatility to be driven not only by the interest rate but also by other economic factors. We illustrate our methodology by applying it to interest rates and other economic factors of the American economy.*

Keywords Interest rates; Markov Chain Monte Carlo; Reversible jump; Threshold model.

JEL Classification C1; C11; C15.

1. INTRODUCTION

There is considerable dispute in econometrics literature on the way that interest rates series should be modelled. The usual phenomenon of volatility clustering is present but, unlike other financial time series, it is accompanied by the special characteristic of positive correlation of volatility and interest rate levels. Modelling the level-volatility relationship has been the major issue in all modelling approaches, which vary across GARCH-type, Markov-switching, and regime-switching models. Our purpose is to introduce a class of flexible regime switching models that can be used to better capture the volatility behavior of interest rates. We shall motivate

Received September 9, 2005; Accepted June 30, 2006
Address correspondence to Petros Dellaportas, Department of Statistics, Athens University of Economics and Business, 76 Patission Str., 10434 Athens, Greece; E-mail: petros@aueb.gr

our modelling approach, firstly, through a careful discussion of existing approaches and then by the illustration of stylized facts.

The most popular models have been developed by continuous time financial theory because they are crucial in the pricing of bond derivatives and in hedging interest rate risk. A general model framework for these models has been developed by Chan et al. (1992) and is represented by the stochastic differential equation

$$dr_t = \{\alpha_0 + \alpha_1 r_t\}dt + \sigma r_t^{\delta} dZ_t, \qquad t \geq 0 \tag{1}$$

where $\{r_t, t \geq 0\}$ is a real continuous time random process representing the interest rate level, α_0, α_1, and δ are parameters, and $\{Z_t, t \geq 0\}$ is a standard Brownian motion. Chan et al. (1992) demonstrate that a series of popular models are in fact special cases of (1) for different values of the parameters, and that a key issue in the above reparameterization is the value of δ which determines the dependency between volatility and interest-rate level. As δ increases, the volatility becomes more sensitive to interest-rate level. Chan et al. empirically conclude that incorporating the parameter δ in the model, rather than taking δ as constant, substantially improves the fit of the model.

The corresponding Euler–Maruyama discrete time approximation of (1) is given by

$$\Delta r_t \equiv r_t - r_{t-1} = \alpha_0 + \alpha_1 r_{t-1} + \epsilon_t, \qquad \epsilon_t \mid \Phi_{t-1} \sim N\left(0, \sigma_t^2 r_{t-1}^{2\delta}\right) \tag{2}$$

where Φ_t is the information set at time t and σ_t^2 is the conditional variance of interest-rate changes.

It is clear that $\delta = 0$ results in models with no level-volatility link, and in the continuous time world (1) are represented by the early models of Merton (1973) and Vasicek (1977) whereas in the discrete time world (2) are expressed through the GARCH-type models (see, for example, Engle et al., 1987) and the Markov-switching models (see, for example, Hamilton, 1988). However, it became evident that the volatility of the short rates is not solely dependent on the rate level. Stylized facts (Lamoureaux and Lastrapes, 1990) demonstrated the need for models that could also capture some form of volatility persistence or clustering. This led to a series of more sophisticated models that incorporated regime switches together with volatility persistence (see Brenner et al., 1996; Chung and Hung, 2000; Dueker, 1997; Gray, 1996). Regime switches in the volatility process could be interpreted as volatility "shocks" occurring at unknown times. This approach has also led to the modelling of the volatility process via Levy processes (for a review, see Barndorff-Nielsen and Shephard, 2001).

There has recently been some effort to explore more complex nonlinear dynamics in the time series of short-term interest rates. This fact has been motivated by the apparent inadequacy of the popular models to capture the peculiar behavior of the data. For example, Ait-Sahalia (1996) suggests replacing the linear drift in (1) by a nonlinear function $\alpha_0 + \alpha_1 r_t + \alpha_2 r^2 + \alpha_3/r$. Another approach which will be the focus of our paper has been suggested by Pfann et al. (1996), who explore autoregressive models with regime switching dynamics that can handle the time-varying persistence of shocks. Further, Lubrano (2000) suggested using nonparametric methods to extend these models.

Pfann et al. (1996) suggested a series of models that are extensions of the threshold models suggested by Tong (1983). A general formulation of these models based on J regimes is

$$\begin{aligned}&\Delta\{g(r_t)\} = \alpha_{j0} + \alpha_{j1} r_{t-1} + \epsilon_t,\\ &\epsilon_t \mid \Phi_{t-1} \sim N(0, \sigma_t^2 r_{t-1}^{2\delta}), \qquad \text{if } r_{t-1} \in [c_{j-1}, c_j], \quad j = 1, \ldots, J\end{aligned} \tag{3}$$

with $g(r_t)$ taken to be either the identity or the log function and Φ_t defined as in (2). Note that the number of regimes J is chosen before the analysis takes place. The difference between these models and Markov switching ones is that the regime switching depends on the level of interest rates, not just time. There is a series of empirical evidence (Lubrano, 2000; Pagan et al., 1994; Pfann et al., 1996) to suggest that, at least for the short-term interest rates in the United States, this is an important stylized fact that should be taken into account.

In this paper we focus on nonlinear threshold models of the types suggested by Pfann et al. (1996) and we offer some modelling alternatives. We relax the assumption of the fixed number of regimes, allowing for automatic model determination through the reversible jump algorithm (Green, 1995). Thus, we extend the previous work in the area of threshold models by jointly estimating both the number of thresholds required as well as the position of these thresholds. This prevents the modeller from making errors by allowing too few, or perhaps even too many, thresholds. Finally, we also allow thresholds in the volatility to be driven, not only by the interest rate, but also by other economic factors. In this way the model attempts to explain possible jumps in the volatility, in contrast to the models that assume these jumps to be realizations of a random process. This could have significant benefits in understanding the relationship between economic factors and the volatility of the interest rates. Other related work in the area of Bayesian threshold models is given by Koop and Potter (1999a,b). See also the papers by Koop and Potter (2004a,b), Kristensen (2004), and Pesaran et al. (2004) that address the issue of structural break models.

Our paper is organized as follows. In Section 2, we describe the data we analyze. In Section 3, we present the existing and the proposed models. In Section 4, we provide guidelines for the Bayesian inference implementation, with particular emphasis on the reversible jump Markov Chain Monte Carlo (MCMC) algorithm used for the case of the unknown number of regimes. Section 5 contains the result of our data analysis exercise and Section 6 concludes with a brief discussion.

2. THE DATA

We focus on the monthly U.S. 3-month treasury bill rates from January 1962 until December 1999, using this as a measure of the short-term interest rate. We shall denote this time series by r_t ($t = 1, \ldots, 456$). This is plotted in Figure 1 together with the long-term rate over the same period.

In Figure 2 we display scatterplots of both the long- and short-term interest rate against the differenced short rate, $\Delta r_t = r_t - r_{t-1}$. Both these plots show how the volatility in Δr_t increases as the magnitude of the rates increases. Let us focus attention first on Figure 2(a). We see that for very low short-term rates (less than around 5%) there is a cluster of datapoints with very little volatility. With rates between 5% and 10% another regime with medium volatility seems to be present. Then, when rates are above 10% there is a great deal of volatility in the rate. These inferences can be seen by "eyeballing" the scatterplot but there remains uncertainty in whether all these regimes can capture the underlying volatility, or if perhaps more (or even less) would provide a better model.

In Figure 3 we plot the differenced short-term rate against the differenced inflation rate. This scatterplot suggests that large drops and large rises in inflation rate from one month to the next increases interest-rate volatility as one would expect (volatility in interest and inflation rate

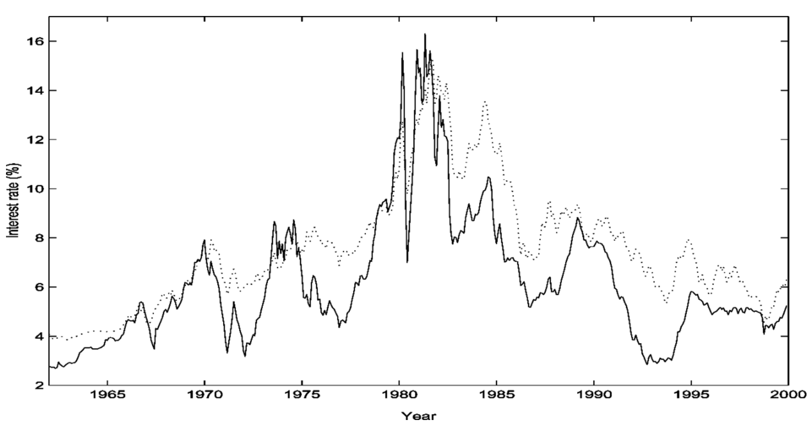

FIGURE 1 The long- and short-term monthly interest rates for the period 1962 to 1999. The solid line is the 3-month treasury bill rate while the dotted line gives the 10-year treasury yield.

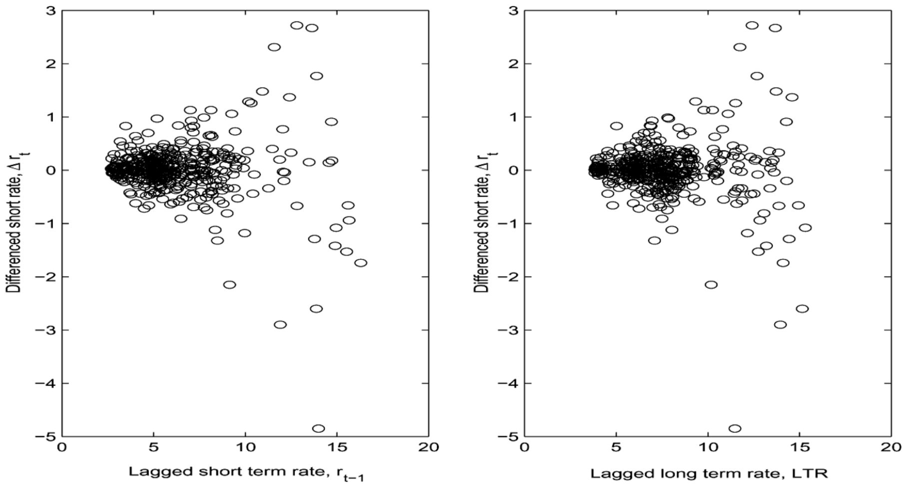

FIGURE 2 Scatterplots of the differences in 3-month treasury bill rate (Δr_t) against lagged values of: (a) 3-month treasury bill rate (r_{t-1}) and (b) the 10-year treasury yield (LTR).

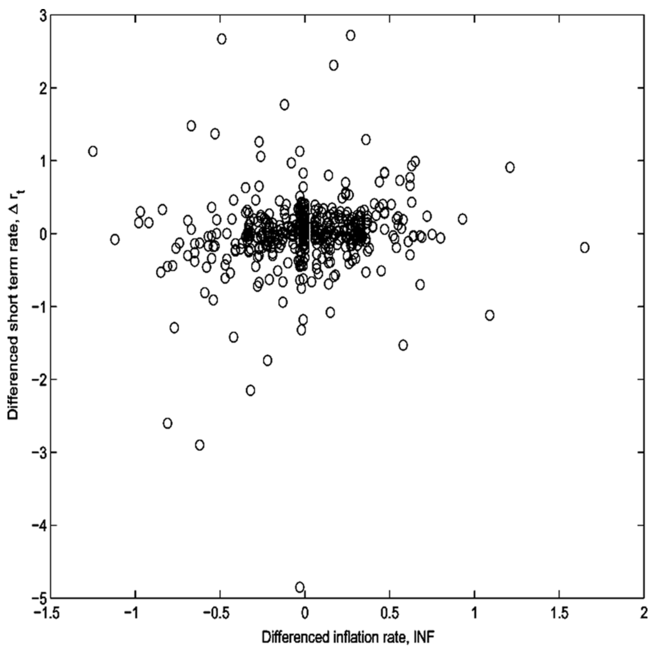

FIGURE 3 Scatterplot of the differences in 10-year treasury yield and the differenced inflation rate. Note how when the inflation rate is low (below about $-.5$) it appears that the variance of Δr_t increases.

occur together). However, there is some evidence of large drops having more effect than large rises as the cluster of points having similar variance in Δr_t looks to be from about $-.3$ to $.5$. This suggests that a nonlinear model is needed to accurately model the inherent volatility.

Motivated by the discussion of Pfann et al. (1996), we shall also make reference to some other economic indicators and we shall test whether they could justify the many episodes of mean-reverting tendencies that appear in interest rate levels. Although such a model would be less parsimonious, it could serve as a guide for predictions based on related covariates; a related application could be estimation of bond prices as in Kristensen (2004). The six series we chose are given in Table 1 with seasonal adjustment having taking place where applicable. A brief description of these indicators follows.

The trade-weighted measure of the US dollar against other major currencies (CUR) was thought to be an important indicator of interest-rate behavior due to its apparent immediate connection with inflation (INF). The producer price index (PPI) estimates the prices received by domestic producers of goods and tends to anticipate prices at the retail level. Its variability reflects, in some sense, the growth or slowdown of American economy and possibly inflation and interest rates. The National association of purchasing management index (NAPM) is a manufacturing index that is published monthly and it is important because it is the first indicator of the U.S. manufacturing economy performance during the previous month. Its behaviour is a possible trend indicator for manufacturing activity and thus a signal for economic growth or slowdown, which in turn may affect inflation and consequently interest-rate behavior. The consumer price index (CPI) measures the rise or fall in prices that consumers pay for a "market basket" of goods. It is a very important indicator for inflation and thus for interest rates. Finally, the monthly U.S. 10-year treasury yield interest rate was included as a measure of the long-term interest rate (LTR). Note that all the series of explanatory variables were differenced before the analysis apart from LTR and the data was retrieved from DATASTREAM.

TABLE 1 The time series used which could affect the long-term US interest rate where $X_i = (x_{i,0}, \dots, x_{i,T-1})$ $(i = 1, \dots, 6)$

INF	US annual inflation rate
CUR	A trade-weighted measure of the US$ against other major currencies
PPI	US Producer Price Index (PPI) for finished goods
NAPM	The National Association of Purchasing Management index (NAPM)
CPI	The US Consumer Price Index (CPI)
LTR	10-year US Treasury yield (measure of the long-term interest rate)

3. MODELS SPECIFICATION

Pfann et al. (1996) analyze the volatility of U.S. interest rates using a self-exciting threshold autoregressive (SETAR) model with two regimes. This model can be written as

$$\Delta r_t = \begin{cases} \alpha_{10} + \alpha_{11} r_{t-1} + \sigma_1 \epsilon_t & r_{t-1} < c \\ \alpha_{20} + \alpha_{21} r_{t-1} + \sigma_2 \epsilon_t & r_{t-1} \geq c \end{cases} \tag{4}$$

where c is the threshold at which there is a regime shift. It simply states that when the interest rate crosses a level c, a model with different parameters is fitted. Throughout this paper we shall adopt a Bayesian approach to inference. For this simple model, we show, in the next section how we can analytically determine the posterior distribution of the threshold value c.

The SETAR model in (4) can be further generalized to the SETAR-TWO one in which the changes in the mean and variance do not necessarily occur together, i.e.,

$$\Delta r_t = \left\{ \begin{array}{ll} \alpha_{10} + \alpha_{11} r_{t-1} & r_{t-1} < c_1 \\ \alpha_{20} + \alpha_{21} r_{t-1} & r_{t-1} \geq c_1 \end{array} \right\} + \left\{ \begin{array}{ll} \sigma_1 \epsilon_t & r_{t-1} < c_2 \\ \sigma_2 \epsilon_t & r_{t-1} \geq c_2 \end{array} \right\}. \tag{5}$$

Instead of assuming that the mean level is subject to regime shifts, we wish to only allow jumps in the volatility. This allows us to estimate coefficient parameters that remain fixed over time, with departures from this model only allowed through jumps in the volatility process. This is worthwhile as it often appears that regime shifts in the SETAR model are driven by volatility jumps, not changes in the underlying mean-level process. Also, we can provide more precise coefficient estimates using all the data. In particular, we allow regime switches in the volatility of r_t using any of the time series listed in Table 1. It is difficult to know how many such regime shifts need to be incorporated in the model before the analysis takes place so we use the data to determine this at the same time as estimating the parameters. We also allow the data to dictate which series are split on. Hence, if $\boldsymbol{x}_{t-1}$ represents the information available at time $t-1$ we take

$$\begin{aligned} \Delta r_t &= \alpha_0 + \alpha_1 r_{t-1} + \sigma_t(\boldsymbol{x}_{t-1})\epsilon_t, \qquad (t = 1, \ldots, T) \\ \sigma_t^2(\boldsymbol{x}_{t-1}) &= \sigma^2\left(1 + \sum_{j=1}^{J} \gamma_j \mathcal{I}_j(\boldsymbol{x}_{t-1})\right), \end{aligned} \tag{6}$$

where J is the number of thresholds in the volatility function and γ_j gives the size of the jump associated with the jth threshold function $\mathcal{I}_j(\boldsymbol{x}_{t-1}) \in \{0, 1\}$ which depends solely on the information available at the previous time point, $\boldsymbol{x}_{t-1}$. Note that γ_js are constrained so that the resulting variance $\sigma_t^2(\boldsymbol{x}_{t-1})$ is positive. The other parameter σ^2 can be thought as the global (static) volatility of r_t that is unchanged by the information available. Of course the parametric specification of the mean in (6) could be more general, see for example Table 1 of Kristensen (2004), or, more generally, it could also include a threshold-based form similar to the one used in the volatility. Whereas the former is a straightforward extension of our methodology, the latter involves a series of complicated identifiability issues that should be taken into account by careful choice of prior formulation; this issue is not studied here.

A possible choice for the threshold functions, and the most popular in the literature, is

$$\mathcal{I}_j(\boldsymbol{x}_{t-1}) = [s_j(x_{v_j,t-1} - c_j)], \quad (j = 1, \dots, J) \tag{7}$$

where $[a] = 1$ if $a > 0$ and 0 otherwise, v_j denotes an explanatory variable which may affect volatility and it is the variable on which $\mathcal{I}_j$ splits, c_j is the threshold point, and $s_j \in \{-1, 1\}$ is required so that $\mathcal{I}_j$ can be nonzero either when $x_{v_j,t-1} > c_j$(with $s_j = 1$) or when $x_{v_j,t-1} \leq c_j$ ($s_j = -1$).

Although threshold functions as in (7) only allow splits dependent on a single component of $\boldsymbol{x}_{t-1}$ they are usually favored over more elaborate threshold functions as they are easier to interpret. This is because models which provide insight into the underlying economics such as ones that could suggest "there is a jump in the volatility function when the lagged interest rate is above 7.5%" are highly valued.

4. POSTERIOR INFERENCE

This section describes how we make inference for the models described in Section 3. As mentioned before, we adopt a Bayesian approach throughout which implies that if $\boldsymbol{\theta}$ is a vector of parameters that describes the model we are fitting, all required inference is based on the posterior distribution of $\boldsymbol{\theta}$. Given data $\mathcal{D}$ this is, from Bayes' Theorem,

$$p(\boldsymbol{\theta} \mid \mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}). \tag{8}$$

where $p(\mathcal{D} \mid \boldsymbol{\theta})$ is the likelihood of the parameters and $p(\boldsymbol{\theta})$ is the prior distribution. Hence Bayesian inference involves assigning a prior to the unknown parameters and updating them through the standard likelihood to form the posterior distribution $p(\boldsymbol{\theta} \mid \mathcal{D})$.

Let us now focus only on the SETAR model of Pfann et al. (1996) and the SETAR-J model we propose. Thus, if we take $\boldsymbol{\theta}$ to be equal to all the unknown parameters in the either model, we find that for the SETAR model

$$\boldsymbol{\theta} = \{\alpha_{10}, \alpha_{11}, \alpha_{20}, \alpha_{21}, \sigma_1, \sigma_2, c\},$$

and for the SETAR-J one

$$\boldsymbol{\theta} = \{\alpha_{10}, \alpha_{11}, \sigma, (\gamma_j, s_j, v_j, c_j)_1^J\}.$$

A Bayesian approach then proceeds by placing prior distributions to all these parameters, and then making inference after observing the data through the posterior distribution defined by (8).

4.1. Priors

We now outline the priors to choose for the parameters, all the time trying throughout to make them as uninformative as possible.

For variances in the two models $\sigma^2, \sigma_1^2, \sigma_2^2$ we assign the conjugate inverse-gamma prior so that for σ^2, $p(\sigma^2) \propto \sigma^{-2}$ and similarly so σ_1^2, σ_2^2. For each of the coefficients we again use the conjugate normal specification with infinite variance so that $p(\alpha_{ij}) \propto 1$. Both these priors are improper (i.e., cannot be normalized) but it is well-known that they lead to proper posterior distributions when using the Bayesian linear model. For the discrete unknown parameters we place discrete uniform (DU) prior distributions on them. That is, the prior on c_j is discrete uniform on the distinct observed values in the dataset, $s_j \sim DU\{-1, 1\}$, $v_j \sim DU\{1, \ldots, p\}$, and $J \sim DU\{1, \ldots, J_{\max}\}$ where $J_{\max}$ is some prespecified maximum number of thresholds. The final prior we need to set is over the γ_js. We take these to be distributed as standard normals a priori as we expect, with great certainty, that they lie between -2 and 2.

We note that this prior parameterization removes the need to specify prior distributions over an increasing number of parameters; for a detailed study of this issue, see Koop and Potter (2004a).

4.2. Likelihood

When analyzing monthly data we can assume that the errors associated with each return are conditionally independent and normally distributed so that $r_t \sim N(\mu_t, \sigma_t^2)$ where μ_t and σ_t depend on some of the other parameters in $\boldsymbol{\theta}$ (such as the thresholds) as well as the lagged return. Hence the log-likelihood is given by

$$\log p(\mathcal{D} \mid \boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=1}^{T} \left[\log(2\pi\sigma_t^2) + \left\{ \frac{\Delta r_t - \mu_t}{\sigma_t} \right\}^2 \right] \qquad \text{for all } \sigma_t > 0. \quad (9)$$

4.3. Inference for the SETAR Model

We note that $\boldsymbol{\theta}$ is seven-dimensional for the SETAR model. Therefore, to determine the target posterior, $p(\boldsymbol{\theta} \mid \mathscr{D})$, we need to perform a numerical approximation to a seven-dimensional integral, $p(\mathscr{D}) = \int p(\mathscr{D} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$. This integral is not available analytically and standard numerical analysis techniques for performing such an integral would not be easy to implement. However, as we have adopted a conjugate prior specification for the α_{ij} and the σ_i we find that we can perform the integration over these parameters analytically.

First note that the SETAR model can be written as a linear model, that is,

$$\boldsymbol{R} = \boldsymbol{X\alpha} + \boldsymbol{\epsilon},$$

where $\boldsymbol{R} = (\Delta r_1, \ldots, \Delta r_T)$, $\boldsymbol{\alpha} = (\alpha_{10}, \alpha_{11}, \alpha_{20}, \alpha_{21})$, $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{X}$ is the design matrix with row t given by $(1\ r_{t-1}\ 0\ 0)$ if $r_{t-1} \leq c$ and by $(0\ 0\ 1\ r_{t-1})$ otherwise. The variance matrix of the errors, $\boldsymbol{\Sigma}$, is diagonal with the tth diagonal element given by σ_1^2 if $r_{t-1} \leq c$ and σ_2^2 otherwise. Thus, this model reduces to two independent linear regressions.

Using standard results (e.g., O'Hagan, 1994, Chapter 9), we find that the log-likelihood of the SETAR model, marginalized over the coefficients $\boldsymbol{\alpha}$ and the error variances σ_1, σ_2, is given by

$$\log p(\mathscr{D} \mid c) = -\frac{1}{2}\sum_{i=1}^{2}\{\log|\boldsymbol{X}_i'\boldsymbol{X}_i| + T_i \log(\boldsymbol{R}_i'\boldsymbol{R}_i - \hat{\boldsymbol{\alpha}}_i\boldsymbol{X}_i'\boldsymbol{X}_i\hat{\boldsymbol{\alpha}}_i)\}, \tag{10}$$

where the i subscript indexes the matrices/vectors relating to those points in the first and second regime. Further, $\hat{\boldsymbol{\alpha}}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{R}_i$, the least squares estimate to α_i and $|\cdot|$ represents the determinant of a matrix. The marginalization of most of the parameters in $\boldsymbol{\theta}$ allows us to write down the posterior distribution of c simply as just

$$p(c \mid \mathscr{D}) = \frac{p(\mathscr{D} \mid c)}{\sum_{c \in \mathscr{C}} p(\mathscr{D} \mid c)}, \tag{11}$$

where $\mathscr{C}_1$ is the set of distinct values of r_t given in the dataset, and the prior term $p(c)$ has been cancelled from the expression as it is a constant that appears in both the numerator and denominator. As this calculation can be performed easily (the summation in the denominator turns out to only contain 334 terms) we can determine the posterior distribution of the changepoint for this model exactly and we display it in Figure 4. This is in contrast to the results of Pfann et al. (1996) who used simulation results for the same problem. These are subject to sampling errors and are less

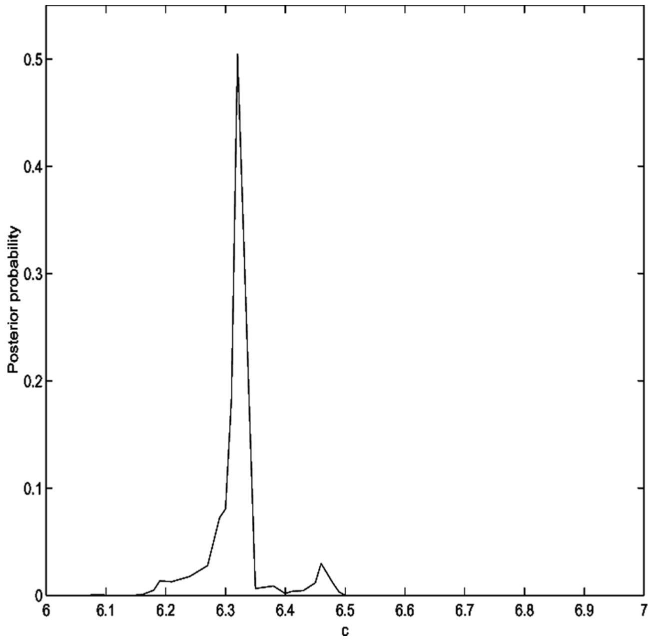

FIGURE 4 The exact posterior distribution for the location of the changepoint, c, for the SETAR model.

computationally efficient. However, for the SETAR-2 model we find that analytic results would not be feasible so that simulation, as demonstrated by Pfann et al. (1996) is appropriate in this case.

4.4. Inference for the SETAR-J Model

In the previous subsection we showed that analytic integration can be used to marginalize out over the coefficients and the variance in the SETAR model. This is also the case for the more complex SETAR-J model and we find that

$$p(\mathscr{D} \mid \{\gamma_j, s_j, v_j, c_j\}_1^J) = -\frac{1}{2}\{|\boldsymbol{\Sigma}| + \log|\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i| + T\log(\boldsymbol{R}'\boldsymbol{\Sigma}^{-1}\boldsymbol{R} - \hat{\boldsymbol{\alpha}}'\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\hat{\boldsymbol{\alpha}}\}, \tag{12}$$

where $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_T^2)$. The major difference between this equation and (10) is that the marginal likelihood for this model depends on $4J$ parameters rather than the single one, c, for the SETAR model. Hence the posterior for the unknown parameters cannot be determined by a tractable summation as in (11) as it involves a repeated summation over all the $4J$ parameters. This is why we must resort to the sampling algorithm that we outline in the next subsection to draw from the posterior of unknown model parameters.

4.5. The Reversible Jump Sampling Algorithm

Calculating the posterior distribution of $\boldsymbol{\theta}$ for the SETAR-J model is more complex than that for the previous models. The two difficulties are the varying dimension of the parameter space and the large number of parameters for most of these models (i.e., $(3+4J)$ parameters for the model with J thresholds). Although analytic integration of the coefficients and the overall regression variance σ^2 can still take place, to make inference over the remaining $4J$ parameters is not straightforward. We choose to make inference by simulating draws from the posterior of interest and then using the obtained generated sample to approximate the true posterior. When the dimension of the space is varying the reversible jump MCMC method of Green (1995) gives us a way to construct an algorithm to do this.

If we assume that the maximum number of thresholds we permit is given by $J_{\max}$ and let a general vector of parameters that describes a model with J thresholds be given by $\boldsymbol{\theta}_J \in \boldsymbol{\Theta}_J$, where $\boldsymbol{\Theta}_J$ is the fixed dimensional subspace containing all the models with J thresholds. So, we find similarly to the derivation of (11), that the posterior probability of J thresholds is given by

$$p(J \mid \mathscr{D}) = \frac{p(J)\int_{\boldsymbol{\Theta}_J} p(\mathscr{D} \mid \boldsymbol{\theta}_J, J)p(\boldsymbol{\theta}_J \mid J)\,d\boldsymbol{\theta}_J}{\sum_{J=1}^{J_{\max}} p(J)\{\int_{\boldsymbol{\Theta}_J} p(\mathscr{D} \mid \boldsymbol{\theta}_J, J)p(\boldsymbol{\theta}_J \mid J)\,d\boldsymbol{\theta}_J\}} \tag{13}$$

where $p(J)$ is the prior probability for the model with J thresholds and $p(\boldsymbol{\theta}_J \mid J)$ is the prior probability of the model parameters, given that there are J thresholds.

Green (1995) introduced a reversible jump MCMC strategy for generating from the joint posterior $p(J, \boldsymbol{\theta}_J \mid \mathscr{D})$, based on the standard Metropolis Hastings approach. During reversible jump MCMC sampling, the constructed Markov chain moves within and between models so that the limiting proportion of visits to a given model is the required $p(J \mid \mathscr{D})$ in (13).

In general, suppose that the current state of the Markov chain at time t is $(J, \boldsymbol{\theta}_J)$, where $\boldsymbol{\theta}_J$ has dimension $d(\boldsymbol{\theta}_J)$, and a move is proposed at time $t+1$ to a new model J' with probability $j(J, J')$ and corresponding parameter vector $\boldsymbol{\theta}'_{J'}$. Then, a vector $\boldsymbol{u}$ is generated from a specified proposal density $q(\boldsymbol{u}|\boldsymbol{\theta}_J, J, J')$ and we set $(\boldsymbol{\theta}_{J'}, \boldsymbol{u}') = g_{J,J'}(\boldsymbol{\theta}_J, \boldsymbol{u})$ for a specified invertible function $g_{J,J'}$ such that $g_{J',J} = g^{-1}_{J,J'}$. Note that $d(\boldsymbol{\theta}_J) + d(\boldsymbol{u}) = d(\boldsymbol{\theta}'_{J'}) + d(\boldsymbol{u}')$. Green (1995) showed that if the new move is accepted as the next realization of the Markov chain with probability $a = \min\{1, r\}$ where

$$r = \frac{p(\mathcal{D} \mid \boldsymbol{\theta}_{J'}, J')p(\boldsymbol{\theta}'_{J'} \mid J')p(J')j(J', J)q(\boldsymbol{u}' \mid \boldsymbol{\theta}'_{J'}, J', J)}{p(\mathcal{D} \mid \boldsymbol{\theta}_J, J)p(\boldsymbol{\theta}_J \mid J)p(J)j(J, J')q(\boldsymbol{u} \mid \boldsymbol{\theta}_J, J, J')}|A|, \tag{14}$$

where $A = \partial(\boldsymbol{\theta}'_{J'}, \boldsymbol{u}')/\partial(\boldsymbol{\theta}_J, \boldsymbol{u})$ is the Jacobian of the transformation, then the chain satisfies the conditions of detailed balance and has the required limiting distribution $p(J, \boldsymbol{\theta}_J \mid \mathcal{D})$.

To implement the reversible jump MCMC we need to specify the probability $j(J, J')$ for every proposed move, the proposal distributions $q(\boldsymbol{u} \mid \boldsymbol{\theta}_J, J, J')$, $q(\boldsymbol{u}' \mid \boldsymbol{\theta}_{J'}, J', J)$ and the function $g_{J,J'}$.

The algorithm we use to sample from the posterior consists of a series of move types, with each possible move type attempted with equal probability. The four possible moves we employ are:

ADD	The addition of an extra threshold in the volatility;
DEL	The deletion of a threshold in the volatility;
CHANGE	The change of threshold (c_j and s_j) and size of jump (γ_j) in one of the threshold functions; and
UPDATE	The update of the current value of the size of jump (γ_j)

CHANGE and UPDATE steps do not propose changes in the dimension of the model so are undertaken with standard Metropolis–Hastings steps (Hastings, 1970). For the CHANGE steps the proposal distributions are drawn from the respective priors and for UPDATE from a normal distribution centred on the current value of γ_j with standard deviation $\frac{1}{2}$. Note that the proposed values of γ_j that result to negative variance $\sigma_t^2(\boldsymbol{x}_{t-1})$ in (6) are immediately rejected.

Now we consider the move types ADD and DEL which propose changes to the dimension of the model. We choose to use straightforward proposal distributions for these changes to simplify expresssion (14). This is done to make the method easier to understand and to reduce the chances of introducing errors due to complex mathematical calculations.

First we shall look at the acceptance probability of the ADD step from current configuration $\boldsymbol{\theta}_J$, with J threshold functions, to proposed configuration $\boldsymbol{\theta}'_{J'}$, with $J' = J + 1$ threshold functions (where we take $\boldsymbol{\theta}$ to only include the model parameters that cannot be analytically integrated out).

From the prior distributions outlined earlier we know that

$$p(J) = p(J') = \frac{1}{J_{\max} + 1}, \tag{15}$$

for $J = 0, 1, \ldots, J_{\max}$, as each number of thresholds is assumed equally likely. Also,

$$p(\boldsymbol{\theta}_J \mid J) = \frac{1}{J!}\frac{1}{2p}\prod_{j=1}^{J}\left\{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\gamma_j^2}{2}\right)\right\} \tag{16}$$

where the $J!$ terms take into account the fact that the model is identical no matter how many ways the thresholds are indexed (and there are $J!$ possible permutations of these indices). A similar expression is found for $p(\boldsymbol{\theta}_{J'} \mid J')$.

When we perform an ADD step we draw the new threshold parameters, $s_{J'}, v_{J'}, \gamma_{J'}$, from their respective prior distributions so that

$$q(\boldsymbol{u} \mid \boldsymbol{\theta}_J, J, J') = \frac{1}{2p}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\gamma_{J'}^2}{2}\right). \tag{17}$$

However, when performing a DEL step we only need to draw the index of the threshold to remove so we find that

$$q(\boldsymbol{u}' \mid \boldsymbol{\theta}_{J'}, J', J) = \frac{1}{J'}. \tag{18}$$

We now define the probabilities of proposing ADD and DEL steps. Again, with simplicity in mind, we propose each of the possible move types for any current value of J with equal probability so:

$$j(J,J')=\frac{1}{4},\quad J=1,\ldots,J_{\max}-1;\quad j(J,J')=1,\quad J=0;\quad j(J,J')=0,\quad J=J_{\max}; \tag{19}$$

$$j(J',J)=\frac{1}{4},\quad J'=1,\ldots,J_{\max}-1;\quad j(J',J)=0;\quad J'=0, j(J',J)=\frac{1}{3},\quad J'=J_{\max}. \tag{20}$$

The last term in (14) that we need to determine is the Jacobian of the transformation between the variables. This turns out to be unity as the new parameters are drawn indepedently of the current ones.

Now, by combining (14)–(20), and after suitable cancellation, we find that the acceptance probability is given simply by

$$a = \min\left\{1, \frac{p(\mathcal{D} \mid \boldsymbol{\theta}'_{J'}, J')}{p(\mathcal{D} \mid \boldsymbol{\theta}_J, J)}\frac{j(J',J)}{j(J,J')}\right\},$$

where the fraction involving the move probabilities, in general, can also be cancelled out as it is unity. A similar algorithm which leads to the same acceptance probability is outlined in more detail in Denison et al. (1998).

5. RESULTS FOR THE SETAR-*J* MODEL

Throughout this paper we show results based on a sample of 10,000 draws from the posterior. These were found by first running the algorithm for 10,000 burn-in iterations, after which time convergence was assumed, and then taking draws every 50 iterations to be in the approximate posterior sample. We also took $J_{\max}$ as 20. No accepted method of testing for non convergence of the Markov chain for such varying-dimension models is available. However, we monitored the autocorrelations in the number of thresholds in the sample to ensure that adequate mixing had taken place. Other usual empirical tests of convergence including graphical inspection of posterior densities had also taken place. In Figure 6 we display the posterior distribution for the number of thresholds and see that many more than are commonly used (two or possible three regimes) are suggested by our model.

In Figure 5 we plot the mean volatility found over the 10,000 models in the sample for the SETAR-J model. This captures the perceived volatility

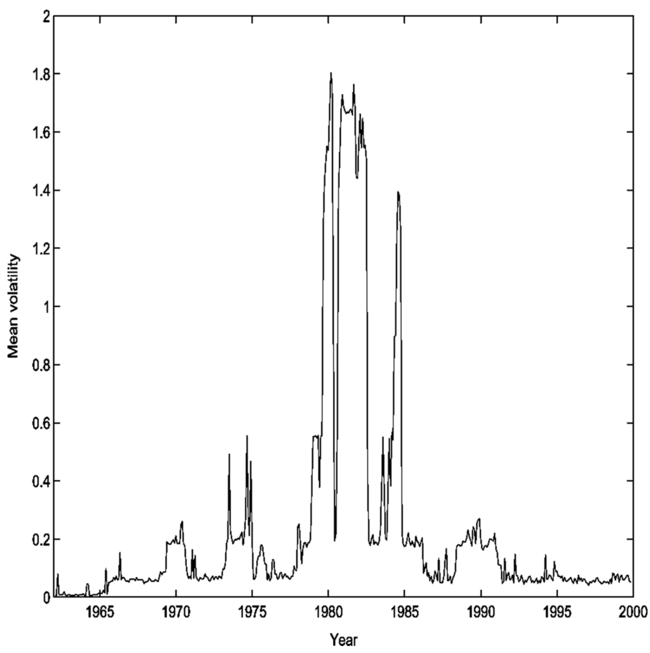

FIGURE 5 Mean volatility given by the model allowing jumps in volatility and only regressing on lagged short-term rate.

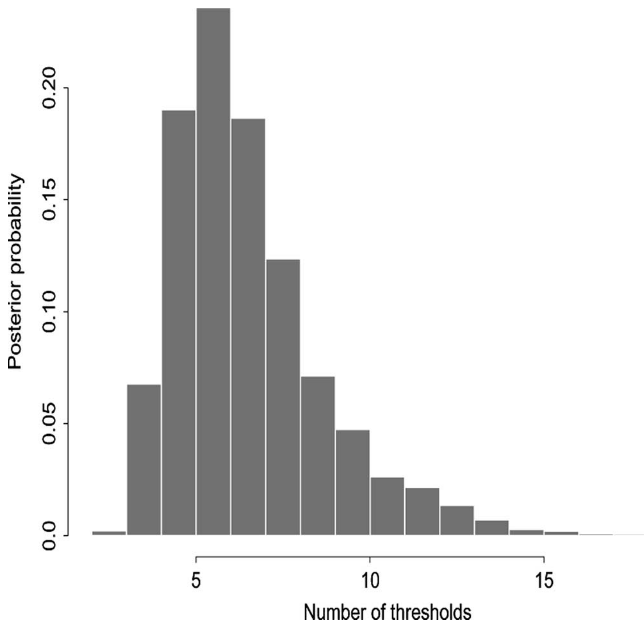

FIGURE 6 The posterior distribution for the number of thresholds.

we can see in Figure 2. The high volatility period between 1979 and 1985 is captured well and demonstrates how the model allows large jumps in the volatility in certain regions without overly affecting the estimated volatility near these periods of high volatility. In the same way, spurious short-lived jumps in volatility can be captured easily (e.g., the period 1973 to the end of 1974 contains two such jumps).

We also want to determine how each variable affects the mean volatility. As each threshold corresponds to only one variable we can plot the threshold functions for each variable, averaged over the generated sample. We do this in Figure 7. Note that these curves are averages over step functions so are by nature discontinuous. However, the posterior averaging does make the resulting mean curves smoother than would be expected if we were just displaying the results from a single model.

The top plot in Figure 7 shows the effect that of the lagged short-term rate, r_{t-1} on the volatility at time t, σ_t^2. We see that there is no effect for short rates less than 6%, then a small effect for rates between 6–9%. After this the effect increases monotonically, but probably not linearly, until we reach high rates of around 13%. Note that the y-scale on this curve is different from the other six curves. These are all plotted on the same scale for ease of comparison. These show that there is little effect on the volatility for all the other variables except possibly, NAPM. Large negative values of this variable (recall that we use NAPM differenced) imply volatility

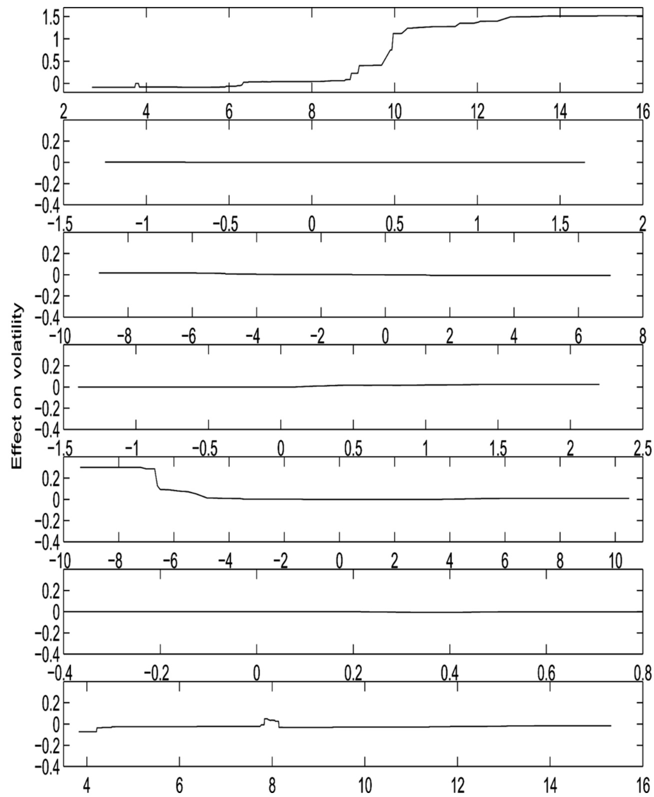

FIGURE 7 Posterior mean effect of each variable on the volatility. From top to bottom: r_{t-1}, INF, CUR, PPI, NAPM, CPI, and LTR.

increases. This can be interpreted as revealing that when there is a signal of economic growth slowdown, the volatility of the interest rate increases.

In Figure 8 we plot the short-term rate against the posterior mean volatility shown in Figure 5. This highlights the nonlinear relationship between volatility and the rate and, to some degree, the shape of the scatterplot mimics the top plot in Figure 7. Again we see evidence of volatility thresholds and notice different regimes in the regions less than 6.5%, between 6.5% and 9%, and a nonlinear relationship for rates greater than about 9%.

6. DISCUSSION

We have extended previous work in the area of SETAR modelling by providing a more general modelling framework with immediate applications in interest rate modelling. It can automatically capture the form of the model by inferring the number, and form, of the threshold functions that can be used to model the volatility of interest rates.

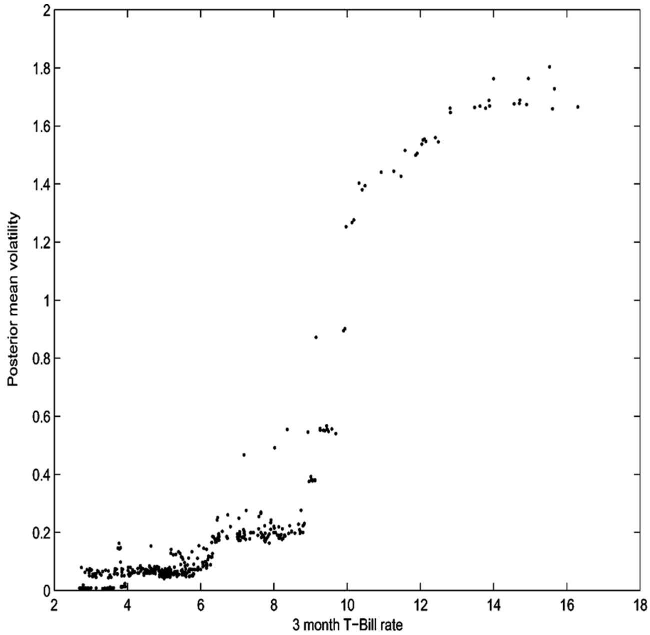

FIGURE 8 Mean volatility against short-term rate.

We used a series of economic indicators to explain volatility jumps. Although only one of these indicators (NAPM) turned out to provide some explanation/predictive power other than the interest rate itself, this exercise served as an illustration for other similar studies. For example, we believe that extensions to stock and exchange rate models will be immediate, and our feedback from discussions with practitioners is that such nonlinear formulations may represent the way volatility traders behave.

A sensible, but not immediate, extension of the models we propose is to model multivariate time series. This has applications in interest rate modelling where short and long interest rates may be modelled together. It will add to the vast literature which examines the simultaneous volatility behavior of stocks in portfolio management and derivative pricing (e.g., Aguilar and West, 2000).

ACKNOWLEDGMENTS

We would like to thank K. Poulopatis and M. Xafa from Citibank, Athens, for useful discussions.

REFERENCES

Aguilar, O., West, M. (2000). Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *J. Bus. Econ. Statist.* 18:338–357.

Ait-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Rev. Financial Studies* 9:385–426.

Barndorff-Nielsen, O. E., Shephard, N. (2001). Non-Gaussian OU based models and some of their uses in financial economics (with discussion). *J. the Royal Statistical Soc. Ser. B* 63:167–241.

Brenner, R. J., Harjes, R. H., Kroner, K. F. (1996). Another look at models of the short-term interest rate. *J. Financial and Quantitative Analysis* 31:85–107.

Chan, K. C., Karolyi, F. A., Longstaff, F. A., Sanders, A. B. (1992). An empirical comparison of alternative models of the short-term interest rates. *J. Finance* 47:1209–1227.

Chung, C.-F., Hung, M.-W. (2000). A general model for short-term interest rates. *Applied Economics* 32:111–121.

Denison, D. G. T., Mallick, B. K., Smith, A. F. M. (1998). Bayesian MARS. *Statistics and Computing* 8:337–346.

Dueker, M. J. (1997). Markov switching in GARCH processes and mean-reverting stock-market volatility. *J. Business and Economic Statistics* 15:26–34.

Engle, R .F., Lilien, D., Robins, R. (1987). Estimating time risk premia in the term structure: the ARCH-*M* model. *Econometrica* 55:391–407.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime switcing process. *J. Financial Economics* 42:27–62.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–732.

Hamilton, J .D. (1988). Rational expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates. *J. Economic Dynamics and Control* 12:385–423.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109.

Koop, G., Potter, S. M. (1999a). Bayes factors and nonlinearity: evidence from economic time series. *J. Econometrics* 88:251–281.

Koop, G., Potter, S. M. (1999b). Dynamic asymmetries in U.S. unemployment. *J. Business and Economic Statistics* 17:298–312.

Koop, G., Potter, S. M. (2004a). Prior elicitation in multiple change-point models. Working Paper No. 04/26, University of Leicester.

Koop, G., Potter, S. M. (2004b). Forecasting and estimating multiple change-point models with an unknown number of change-points. Working Paper No. 04/31, University of Leicester.

Kristensen, D. (2004). A semiparametric single-factor model of the term structure. FMG Discussion Papers DP501, London School of Economics.

Lamoureaux, C. G., Lastrapes, W. D. (1990). Persistence in variance, structural change and the GARCH model. *J. Business and Economic Statistics* 8:225–234.

Lubrano, M. (2000). Bayesian non-linear modellings of the short term US interest rate: the help of non-parametric tools. Techinical report: GREQAM-CNRS, France and CORE, Belgium.

Merton, R. C. (1973). Theory of rational option pricing. *Bell J. Economics and Management Science* 4:141–183.

O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics: Bayesian Inference.* Vol. 2b, UK: Edward Arnold.

Pagan, A. R., Hall, A. D., Martin, V. (1994). Modelling the term structure. In: Maddala, G. S., Rao, C. R., eds. *Handbook of Statistics.* 14, pp. 91–118.

Pfann, G. A., Schotman, P. C., Tschernig, R. (1996). Nonlinear interest rate dynamics and implications for the term structure. *J. Econometrics* 74:1549–176.

Pesaran, H., Pettenuzzo, D., Timmermann, A. (2004). Forecasting time series subject to multiple structural breaks. Working paper.

Tong, H. (1983). *Threshold Models in Non-Linear Time Series Analysis.* New York: Springer-Verlag.

Vasicek, O. (1977). An equilibrium characterisation of the term structure. *J. Financial Economics* 5:177–188.